JOURNAL OF MULTIVARIATE ANALYSIS 30, 245–254 (1989)

Optimality of the Least Squares Estimator

ROBERT BERK* AND JIUNN T. HWANG†

Rutgers University

Communicated by the Editors

In a standard linear model, we explore the optimality of the least squares estimator under assuptions stronger than those for the Gauss–Markov theorem. The conclusion is then much stronger than that of the Gauss–Markov theorem. Specifically, two results are cited below: Under the assumption that the unobserved error ε has a spherically symmetric distribution, the least squares estimator for the regression coefficient β is shown to maximize the probability that $\hat{\beta}-\beta$ stays in any symmetric convex set among linear unbiased estimators $\hat{\beta}$. With the additional assumption that ε is unimodal, the conclusion holds among equivariant estimators. The import of these results for risk functions is also discussed.

1. INTRODUCTION

We study the standard linear model:

$$\underset{n\times 1}{y} = \underset{n\times p}{X}\ \underset{p\times 1}{\beta} + \underset{n\times 1}{\varepsilon}\ .$$

Here y is the observation, X the known design matrix of rank p, β the unknown parameter to be estimated, and ε the unobserved error. Throughout, we assume that ε is spherically distributed in R^n. (The case where ε is centered elliptically, meaning $\varepsilon \sim B\delta$ with δ spherical in R^n and B a known nonsingular matrix, can be reduced to this case by the linear transformation B^{-1}.) The least squares estimator for β is $\hat{\beta}_{\rm LS} = (X'X)^{-1} X'y$. The Gauss–Markov theorem asserts (nontrivially when $E|\varepsilon|^2 < \infty$) that $\hat{\beta}_{\rm LS}$ is the best linear unbiased estimator for β in the sense of minimizing the covariance matrix with respect to positive definiteness. In

Received May 20, 1985; revised January 3, 1989.

Key words and phrases: Gauss–Markov Theorem, unbiased estimator $\hat{\beta}$, spherically symmetric distribution.

AMS 1980 subject classifications: Primary 62J05, 62H12; Secondary 62C99.

* Research supported by National Science Foundation Grant MCS-8301082.

† Research supported by National Science Foundation Grant DMS-8809016.

0047-259X/89 $3.00

683/30/2-6

fact, this is equivalent to asserting that for every matrix A, $\hat{\beta}_{\text{LS}}$ uniformly minimizes the risk function

$$E_\beta |A(\hat{\beta}-\beta)|^2 \tag{1.1}$$

among all linear unbiased estimators $\hat{\beta}$. Assuming that ε has a spherically symmetric distribution, one can strenghten the Gauss–Markov theorem. (For a survey paper on spherically symmetric distributions, see Chmielewski [3].)

Before we discuss existing results for this problem, we introduce another criterion for comparing estimators $\hat{\beta}$, which is based on the concentration probability $P(\hat{\beta}-\beta \in C)$ for some prespecified set C. The set C is usually taken to be a convex set containing the origin. Consequently, the larger the concentration probability, the better the estimator. It is of interest to find most concentrated estimators with respect to sets. (That is, estimators maximizing the concentration probability for all β. Such estimators, of course, exist only when the class of competing estimators is somewhat restricted.)

For the linear model, Ali and Ponnapalli [1], Sinha and Drygas [10], and Hwang [6] independently proved:

THEOREM 1.1. *Let ε be spherical. Then among the class of all linear unbiased estimators, $\hat{\beta}_{\text{LS}}$ is most concentrated with respect to all ellipses symmetric about the origin.*

As pointed out in Theorem 3.1 of Hwang [6], this conclusion is equivalent to the assertion that $\hat{\beta}_{\text{LS}}$ uniformly minimizes the risk functions

$$E_\beta L(|A(\hat{\beta}-\beta)|) \tag{1.2}$$

for any matrix A and any nondecreasing real-valued function L. Since (1.2) is much more general than (1.1), their conclusions are much stronger than that of the Gauss–Markov theorem.

In this paper, we strengthen Theorem 1.1. Specifically, among the same class of estimators, we show that $\hat{\beta}_{\text{LS}}$ is most concentrated with respect to (not only ellipses but also) any convex set symmetric about the origin (see Theorem 2.2). The significance of this generalization can be easily seen by considering risk functions. Our theorem implies that $\hat{\beta}_{\text{LS}}$ uniformly minimizes not only (1.2) but also the more general risk function

$$E_\beta L(g(\hat{\beta}-\beta)). \tag{1.3}$$

(See Theorem 2.3.) In (1.3), L is as before and g is a nonnegative symmetric convex function defined on R^p. Interesting examples of loss functions that can be handled as a result are given after Theorem 2.3.

Furthermore, if one is willing to assume unimodality of ε at the origin, the same conclusion holds for a larger class of estimators (namely, the equivariant estimators). (See Theorem 2.8.) Without the unimodality assumption the latter theorem fails. (Example 2.6 is a counterexample.) Many definitions of multivariate unimodality are discussed in Section 2; all but one are shown in Section 3 to be equivalent for spherically symmetric distributions.

2. Optimality of the Least Squares Estimator

We first discuss the notion of a most concentrated estimator. The least squares estimator $\hat{\beta}_{\mathrm{LS}}$ is called *most concentrated* in the class $\mathscr{D}$ of estimators with respect to the sets $C \in \mathscr{C}$, if for every $\hat{\beta} \in \mathscr{D}$, $C \in \mathscr{C}$, and every β,

$$P_\beta(\hat{\beta}_{\mathrm{LS}} - \beta \in C) \geqslant P_\beta(\hat{\beta} - \beta \in C). \tag{2.1}$$

The relationship between this criterion and risk functions is given next in Lemma 2.1, which generalizes Lemma 2.1 of Hwang [6].

Lemma 2.1. *The following two statements are equivalent:*

(i) $\hat{\beta}_{\mathrm{LS}}$ *is most concentrated with respect to* $\mathscr{C}$ *among the estimators in* $\mathscr{D}$.

(ii) $\hat{\beta}_{\mathrm{LS}}$ *minimizes, among* $\hat{\beta} \in \mathscr{D}$, *the risk function* $E_\beta\, g(\hat{\beta} - \beta)$ *for any function* g *bounded from below whose contours* $C = \{x\colon g(x) \leqslant c\}$ *are in* $\mathscr{C}$ *for every* $c \in R$.

Proof. That (i) implies (ii) follows on noting that under (i), $g(\hat{\beta}_{\mathrm{LS}} - \beta)$ is stochastically no greater than $g(\hat{\beta} - \beta)$, for every $\hat{\beta} \in \mathscr{D}$. Conversely, one can show that (ii) implies (i) by letting g be one minus the indicator function of C. ∎

Theorem 1.1 is next generalized to any symmetric convex set (not necessarily an ellipse):

Theorem 2.2. *Let ε be spherically distributed. Then among the linear unbiased estimators the least squares estimator $\hat{\beta}_{\mathrm{LS}}$ is most concentrated for convex symmetric sets.*

Proof. Write a linear unbiased estimator as $\hat{\beta} = \hat{\beta}_{\mathrm{LS}} + Dy$, where D is a $p \times n$ constant matrix. Unbiasedness of $\hat{\beta}$ implies that DX is a zero matrix. Therefore the covariance matrix of $\hat{\beta}$ is

$$\begin{aligned} \operatorname{cov} \hat{\beta} &= \operatorname{cov} \hat{\beta}_{\mathrm{LS}} + 2\sigma^2 (X'X)^{-1} X'D' + \operatorname{cov}(Dy) \\ &= \operatorname{cov} \hat{\beta}_{\mathrm{LS}} + \operatorname{cov} Dy \end{aligned}$$

which is no smaller than cov $\hat{\beta}_{\text{LS}}$ in the sense of positive definiteness. Since $\hat{\beta}_{\text{LS}}$ has an elliptical distribution, Theorem 3 of Fefferman, Jodeit, and Perlman [5] implies that for any symmetric convex set C,

$$P_\beta(\hat{\beta}_{\text{LS}} - \beta \in C) \geqslant P_\beta(\hat{\beta} - \beta \in C), \qquad \text{for all } \beta. \quad \blacksquare$$

To see the statistical implication of the generalization, consider loss functions of the form $L(g(\hat{\beta} - \beta))$ as in (1.3). Obviously, $L(g(\cdot))$ has symmetric and convex contours and therefore Lemma 2.1 and Theorem 2.2 imply

THEOREM 2.3. *For any linear unbiased estimator* $\hat{\beta}$,

$$E_\beta L(g(\hat{\beta}_{\text{LS}} - \beta)) \leqslant E_\beta L(g(\hat{\beta} - \beta)), \qquad \textit{for all } \beta, \tag{2.2}$$

where L is a nondecreasing function that is bounded from below and g is convex and symmetric about the origin (*i.e.*, $g(t) = g(-t)$).

Some interesting symmetric and convex functions g are:

(i) $g_1(t) = t'Kt$, where K is an arbitrary nonnegative definite matrix;

(ii) $g_2(t) =$ maximum of the absolute coordinates of the vector Kt for any matrix K;

(iii) $g_3(t) = \sum_{i=1}^{p} |k_i t_i|$ for any real numbers k_i.

Since Theorem 1.1 applies only to ellipses, it implies (2.2) for g such that $\{x\colon g(x) \leqslant c\}$ is an ellipse. Only g_1 satisfies this condition. Results concerning other g's such as g_2 and g_3 are therefore new. Obviously domination also holds for a linear combination (with positive weights) of a sequence of these loss functions.

Below we generalize Theorem 2.1 to some extent by considering a larger class of estimators. Professor W. Strawderman, in a verbal communication, posed a question as to whether one can generalize Theorem 2.2 or Theorem 1.1 to the class of unbiased estimators. The answer is negative, even for the normal case.

EXAMPLE 2.4. Take the simplest case where we have one normal observation y with unknown mean β and variance 1. Obviously the least squares estimator is $\hat{\beta}_{\text{LS}} = y$. Consider $\hat{\beta} = vy$, where v is a random variable independent of y having the distribution: $P(v=0) = P(v=2) = \frac{1}{2}$. Hence $\hat{\beta}$ is unbiased. However, $P_0(|\hat{\beta}| \leqslant c) = \frac{1}{2} + \frac{1}{2} P_0(|y| \leqslant c/2)$ which is greater than $P_0(|y| \leqslant c/2)$ if $0 < c < 1.02$. The number 1.02 is obtained by solving the equation $1 - \Phi(a/2) = 2(1 - \Phi(a))$, where Φ is the cumulative distribution of a standard normal. Hence $\hat{\beta}_{\text{LS}}$ is not the most concentrated estimator with respect to the interval $(-c, c)$, $0 < c < 1.02$.

However, it is possible to establish further results for (translation) equivariant estimators, the relevant transformations being $y \to y + Xb$. Note that this will transform β to $\beta + b$. An estimator δ is equivariant if for every $b \in R^p$,

$$\delta(y + Xb) = \delta(y) + b. \tag{2.3}$$

Obviously, the class of equivariant estimators contains (and is much larger than) the class of all linear unbiased estimators. (It is also essentially contained—up to translation—in the class of unbiased estimators.)

In dealing with an equivariant estimator $\delta(y)$, we will need the following representation. Let $A = (X'X)^{-1} X'$, so that $\hat{\beta}_{\text{LS}} = Ay$. Moreover, let $P = XA$ be the projection matrix. Substituting $-\hat{\beta}_{\text{LS}}$ for b in (2.3) gives

$$\delta(y) = \hat{\beta}_{\text{LS}} + \delta((I-P)\, y) = \hat{\beta}_{\text{LS}} + \delta((I-P)\, \varepsilon).$$

The particular case $\beta = 0$ will be of interest in establishing later theorems. For this case $y = \varepsilon$ and $\hat{\beta}_{\text{LS}} = A\varepsilon = AP\varepsilon$, since $A = AP$. Rotating coordinates then allows us to obtain the representation

$$\hat{\beta}_{\text{LS}} = B\varepsilon_1^*, \qquad B \text{ some constant nonsingular matrix,} \tag{2.4}$$

and

$$\delta(y) = \hat{\beta}_{\text{LS}} + h(\varepsilon_2^*), \qquad h \text{ some function from } R^{n-p} \text{ to } R^p, \tag{2.5}$$

where $\binom{\varepsilon_1^*}{\varepsilon_2^*}$ has the same distribution as ε and ε_1^* is p-dimensional.

Now we are ready to prove

THEOREM 2.5. *Suppose that g: $R^p \to R$ is a convex function which is symmetric about the origin and bounded from below. Then for any equivariant estimator $\delta(y)$*

$$E_\beta\, g(\delta(y) - \beta) \geqslant E_\beta\, g(\hat{\beta}_{\text{LS}} - \beta). \tag{2.6}$$

Proof. Due to the fact that both $\hat{\beta}_{\text{LS}}$ and $\delta(Y)$ are equivariant, it is sufficient to show that (2.6) holds for $\beta = 0$. By (2.4) and (2.5), $E_0\, g(\delta(y)) = E_0\, g(B\varepsilon_1 + h(\varepsilon_2))$. Since $\varepsilon' = (\varepsilon_1', \varepsilon_2')$ and $(-\varepsilon_1', \varepsilon_2')$ have the same distribution, the last two expectations equal $E_0\, g(-B\varepsilon_1 + h(\varepsilon_2))$, which in turn equals $E_0\, g(B\varepsilon_1 - h(\varepsilon_2))$, by symmetry of g. Using convexity of g, we obtain

$$\begin{aligned} E_\beta\, g(\hat{\beta}_{\text{LS}} - \beta) &= E_0\, g(B\varepsilon_1) \leqslant \tfrac{1}{2} E_0\, g(B\varepsilon_1 - h(\varepsilon_2)) + \tfrac{1}{2} E_0\, g(B\varepsilon_1 + h(\varepsilon_2)) \\ &= E_0\, g(\delta(y)) = E_\beta\, g(\hat{\beta} - \beta). \quad \blacksquare \end{aligned}$$

This theorem show that $\hat{\beta}_{LS}$ is the Pitman (i.e., the best equivariant) estimator relative to any convex symmetric loss in the spherical case. Although Theorem 2.5 deals with a larger class of estimators than Theorem 2.3, the class of loss functions considered is much smaller. (In fact, the loss function in (2.6) corresponds to the loss in (2.2) with L being the identity function.) Can one enlarge the class of loss functions? For example, is the least squares estimator most concentrated equivariant for spheres (or symmetric convex sets)? The following example indicates that the answer in general is negative.

EXAMPLE 2.6. Assume that y is a 1-dimensional random variable with the p.d.f.

$$\begin{aligned} f(t-\beta) &= \tfrac{1}{2}, \qquad 1<|t-\beta|<2, \\ &= 0, \qquad \text{otherwise.} \end{aligned}$$

Obviously the least squares estimator for β is y. However, $P(|y-\beta| \leqslant 1)=0$. The equivariant estimator $\hat{\beta}=y+1$ will give $P(|\hat{\beta}-\beta| \leqslant 1)=P(-2 \leqslant y-\beta \leqslant 0)=\frac{1}{2}$. Therefore the least squares estimator is less concentrated than $\hat{\beta}$ for the symmetric interval $(-1, 1)$.

This counterexample suggests that some unimodality assumption is needed in order to generalize Theorem 2.5 to a larger class of loss functions. With this assumption, we have the following:

THEOREM 2.7. *Assume that ε is spherically symmetric and unimodal at the origin. Then $\hat{\beta}_{LS}$ is most concentrated (for symmetric convex sets) among all equivariant estimators.*

Before we prove Theorem 2.7, we have to clarify what is meant by "unimodal at the origin" in the multivariate case. There is a generally accepted definition of 1-dimensional unimodality: The distribution F is unimodal at 0, if $F(t)$ is convex for negative t and concave for positive t. However, there is no universally accepted definition of multivariate unimodality. Dharmadhikari and Jogdeo [4] presented various definitions of multivariate unimodality, some of which are listed below. We use "UM" to denote "unimodal at the origin." Note that ε here need not be spherical.

(i) ε is linear UM if $a'\varepsilon$ is UM (at 0) for every $a \in R^n$.

(ii) ε is mixture UM if it is a weak limit of mixtures of uniform distributions on compact symmetric convex sets. (This was called convex UM in [4].)

(iii) ε is probability UM if $P(\varepsilon \in C+kt)$ is nonincreasing in $k \in [0, \infty)$ for every symmetric convex set C and every $t \in R^n$.

Before we present property (iv), we introduce a definition. A random vector ε is said (in Olshen and Savage [8]) to be $\alpha-\mathrm{UM}$ if $s^{\alpha}Eg(s\varepsilon)$ is nondecreasing in $s>0$ for every bounded nonnegative measurable function $g\colon R^n \to R$. We are, however, most interested in the property:

(iv) ε is $n-\mathrm{UM}$.

(v) Assume that ε has a p.d.f. f on $R^n-\{0\}$. ε is said to be convex UM if (a) f is symmetric about the origin (i.e., $f(t)=f(-t)$), and (b) for every $c \in R$ the set $\{t \mid f(t) \geqslant c\}$ is convex. (It is possible that ε has a point mass at the origin.)

Definition (v) was shown in Anderson [2] to imply probability UM.

Reference [4] stated without proof that for spherical distributions, (ii), (iii), and (iv) are all equivalent. In the case that ε has a spherical distribution, we will also discuss the following intuitively appealing definition.

(vi) ε is p.d.f. UM at 0 if the p.d.f. $f(|t|)$ of ε exists on $R^n-\{0\}$ and $f(\cdot)$ is nonincreasing. (It is possible that ε has a point mass at the origin.)

In the next section, we show that definition (i) does not seem to be an appropriate definition. Indeed, we show in Theorem 3.1 that *any* n-dimensional ($n \geqslant 3$) spherical distribution is linear UM. The remaining properties (ii) through (vi) are also shown to be equivalent in the spherical case (Theorem 3.2). Any one of them is therefore used as a definition for UM in Theorem 2.7. Using this interpretation, we now establish

Proof of Theorem 2.7. Let C be a convex set symmetric about 0. For any equivariant estimator $\hat{\beta}$, we want to prove that (2.1) holds. For this, it is enough to consider $\beta=0$. That is, $P_0(\hat{\beta} \in C) \leqslant P_0(\hat{\beta}_{\mathrm{LS}} \in C)$, or by (2.4) and (2.5), $P_0(B\varepsilon_1 + h(\varepsilon_2) \in C) \leqslant P_0(B\varepsilon_1 \in C)$. Thus it suffices to show that

$$P_0(\varepsilon_1 + B^{-1}h(\varepsilon_2) \in B^{-1}C \mid \varepsilon_2) \leqslant P_0(\varepsilon_1 \in B^{-1}C \mid \varepsilon_2), \tag{2.7}$$

where $B^{-1}C=\{B^{-1}x\colon x \in C\}$.

Using definition (v) or (vi) of UM, it is easy to see that the conditional distribution of ε_1 given ε_2 is also UM. Using the fact that $B^{-1}C$ is symmetric and convex, (2.7) follows from definition (iii). ∎

Using Lemma 2.1 and Theorem 2.7, one can establish the following theorem.

THEOREM 2.8. *With assumptions as in Theorem* 2.7, (2.2), *holds for all equivariant estimators* $\hat{\beta}$.

3. The Equivalence of Definitions of Multivariate Unimodality under Sphericity

We first show (Theorem 3.1) that linear UM does not seem to be an appropriate notion for UM. Before we prove Theorem 3.1, we discuss a related representation of Khinchine [7]. In the univariate case, Khinchine [7] showed that ε is UM if and only if

$$\varepsilon \sim ru. \tag{3.1}$$

Here the random variable r is independent of u, a uniform $[0, 1]$ random variable.

Theorem 3.1. *Any n-dimensional spherical distribution is linear* UM *for* $n \geqslant 3$.

Proof. Let ε be an n-dimensional random vector having a spherical distribution. The distribution of $a'\varepsilon$ is (apart from a scale factor $|a|$) that of ε_1, the first coordinate of ε.

Write $\varepsilon = rd$ so that $r = |\varepsilon|$ and the random vector d are independent, where d is uniformly distributed over the (surface of the) unit sphere. (Of course, $d = \varepsilon/|\varepsilon|$ when $\varepsilon \neq 0$. When $\varepsilon = 0$ also $r = 0$, so $\varepsilon = rd$ regardless of the choice of d.) Let d_1 be the first coordinate of d. Now, direct Jacobian transformation shows that

$$d_1^2 \sim B_{1/2, (n-1)/2},$$

where $B_{p,q}$ represents a Beta random variable with parameters p and q. Letting z_1 and z_2 be two independent Gamma random variables having equal scale parameters and respective shape parameters p and q,

$$z_1/(z_1 + z_2) \sim B_{p,q}$$

and is independent of $z_1 + z_2$. Using this fact repeatedly, we see that

$$B_{1/2, (n-1)/2} \sim B_{1/2, 1} B_{3/2, (n-3)/2},$$

where, on the right-hand side of the last expression, the two beta random variables are independent. Hence

$$|d_1| \sim (B_{3/2, (n-3)/2})^{1/2} (B_{1/2, 1})^{1/2},$$

which implies

$$|\varepsilon_1| \sim r(B_{3/2, (n-3)/2})^{1/2} (B_{1/2, 1})^{1/2}.$$

Note that $(B_{1/2,1})^{1/2}$ has a uniform distribution over [0, 1]. By Khinchine's representation, $|\varepsilon_1|$ is UM and therefore so is ε_1. ∎

The last proof obviously does not work when $n \leqslant 2$. In fact for $n=2$, the theorem fails as shown in Example 2.1 of [4].

Next, for spherically symmetric distributions, we establish the equivalence of the other definitions of UM discussed in Section 2. We also consider Khinchine's representation (3.1) in higher dimensions. Here we naturally interpret r to be a scalar random variable independent of the n-variate random vector u uniformly distributed over the (interior of the) unit ball centered at the origin.

It turns out that for a spherically symmetric distribution, (3.1) holds if and only if ε is UM. Indeed, we will show that (3.1) is equivalent to all of the notions of UM considered above, except for linear UM.

To do so, we first argue that under sphericity,

$$\begin{aligned} &\text{convex UM} \equiv \text{p.d.f. UM} \Rightarrow \text{mixture UM} \\ &\qquad \equiv \varepsilon \sim ru \Rightarrow \text{probability UM} \Rightarrow n-\text{UM}. \end{aligned} \tag{3.2}$$

Obviously in the spherical case, convex UM is equivalent to p.d.f. UM. Now any p.d.f. UM distribution can be constructed as a weak limit of mixtures of uniform distributions over (the interior of) balls centered at the origin. These are precisely the distributions of scale mixtures of uniform distributions on balls. Hence ε has a spherical mixture UM distribution if and only if $\varepsilon \sim ru$.

By a theorem of Sherman (also Theorem 3.3 in [4]), mixture UM implies probability UM. This later property implies $n-$UM by Theorem 4.1 in [4]. Hence (3.2) is now established.

Now we are ready to prove

THEOREM 3.2. *For spherical distributions, all the properties in* (3.2) *are equivalent.*

Proof. All we need to show is that if ε is $n-$UM, then ε is p.d.f. UM. We do this first assuming the distribution has a p.d.f. except possibly for some mass at the origin.

As in the proof of Theorem 3.1, let $\varepsilon = rd$, where $r = |\varepsilon|$ and the random vector d are independent and d is uniformly distributed over the (surface of the) unit sphere. By Theorem 4 of Olshen and Savage [8], r is $n-$UM. This means that r^n is $1-$UM, by Lemma 2 of the same paper. Therefore r^n (and hence r) has a p.d.f., except that it might have a point mass at the origin. Hence ε has the same property. Let the p.d.f. of ε on $R^n - \{0\}$ be $f(|t|)$.

To finish the proof, we show that $f(\cdot)$ is nonincreasing. The p.d.f. of

$S = r^n$ (on $(0, \infty)$) is proportional to $f(s^{1/2})$. Since S is $1 - \text{UM}$, its p.d.f. on $(0, \infty)$ is nonincreasing (by, say, Theorem 3 of Olshen and Savage [8]). Therefore $f(\cdot)$ is nonincreasing. ∎

Acknowledgments

The authors thank Professor Arthur Cohen, Michael Perlman, Andrew Rukhin, and William Strawderman for their helpful comments and discussions. The second author also thanks the faculty members of the Statistics Department, Rutgers University, for their hospitality during a 1-week visit there.

References

1. Ali, M. M., and Ponnapalli, R. (1982). An optimum property of Gauss Markov estimates when errors have elliptic distributions. *Inst. Math. Statist. Bull.* Abstract 182, **11**, No. 5 (64) 243.
2. Anderson, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc. Amer. Math. Soc.* **6** 170–176.
3. Chmielewski, M. A. (1981). Elliptically symmetric distributions: A review and bibliography. *Internat. Statist. Rev.* **49** 67–74.
4. Dharmadhikari, S. W., and Jogdeo, K. (1976). Multivariate unimodality. *Ann. Statist.* **4** 607–613.
5. Fefferman, C., Jodeit, M., and Perlman, M. D. (1972). A spherical measure inequality for convex sets. *Proc. Amer. Math. Soc.* **33** 114–119.
6. Hwang, J. T. (1985). Universal domination and stochastic domination—estimation simultaneously under a broad class of loss functions. *Ann. Statist.* **13** 295–314.
7. Khinchine, A. Y. (1938). On unimodal distributions. *Izv. Nauchno-Issled. Inst. Math. Mekh.* **2** 1–7.
8. Olshen, R. A., and Savage, L. J. (1970). Generalized unimodality. *J. Appl. Probab.* **7** 21–34.
9. Sherman, S. (1955). A theorem on convex sets with applications. *Ann. Math. Statist.* **26** 763–767.
10. Sinha, B. K., and Drygas, H. (1985). Robustness in linear models. In *Proceedings First Tampere Sem. Linear Models, Tampere University.*